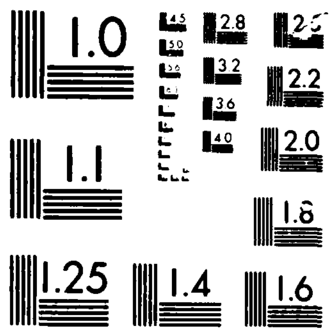
AD-A169 066 A TEST ON THE RELIABILITY AND PERFORMANCE OF THE VERBEX SERIES 4000 VOICE..(U) SYSTEMS CONTROL INC (VT) PALO ALTO CA AERONAUTICAL AND MARINE.. P SUNTHARALINGAM 1/1

UNCLASSIFIED SEP 85 MAE-1748 ARO-20155.6-MA F/G 17/2 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

ARO 20155.6-MA

2

Princeton University

A TEST ON THE RELIABILITY AND PERFORMANCE OF THE VERBEX SERIES 4000 VOICE RECOGNIZER

P. Suntharalingam

September 1985

Technical Report MAE 1748

Department of
Mechanical and
Aerospace Engineering

2

A TEST ON THE RELIABILITY AND PERFORMANCE OF THE VERBEX SERIES 4000 VOICE RECOGNIZER

P. Suntharalingam

September 1985

Technical Report MAE 1748

DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING

PRINCETON UNIVERSITY

PRINCETON, NJ 08544.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 20155.6-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER A169008 N/A
4. TITLE (and Subtitle) A Test on the Reliability and Performance of the Verbex Series 4000 Voice Recognizer		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. Suntharalingam		8. CONTRACT OR GRANT NUMBER(s) DAAG29-84-K-0048
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Princeton, NJ 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September 1985 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Voice recognition systems are becoming increasingly widespread as forms of data entry. One such use of speech input would be as an aid to pilot communication in the cockpit. The Verbex Series 4000 Voice Recognizer (VVR) was chosen as the input channel for a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT CONTINUED

forthcoming flight simulation system. The VVR is a speaker dependent unit with the ability to recognize continuous speech. An additional feature of the VVR is its use of structured grammars in defining the speech format. Tests were run to determine the VVR's reliability, and also to investigate the variations in performance for different grammar structures.

ABSTRACT

Voice recognition systems are becoming increasingly widespread as forms of data entry. One such use of speech input would be as an aid to pilot communication in the cockpit. The Verbex Series 4000 Voice Recognizer (VVR) was chosen as the input channel for a forthcoming flight simulation system. The VVR is a speaker dependent unit with the ability to recognize continuous speech. An additional feature of the VVR is its use of structured grammars in defining the speech format. Tests were run to determine the VVR's reliability, and also to investigate the variations in performance for different grammar structures.

1. INTRODUCTION

Speech input is becoming an increasingly widespread form of data entry. This trend can be attributed to the fact that speech is probably the most natural form of human-machine communication currently available.

It has several advantages over conventional forms of data input such as keyboard entry or operation of a control panel. It needs little training, requiring only that the user confine his speech to a previously defined format. It allows simultaneous communication with humans and machines, and also permits a great deal of mobility and freedom for the user to engage in other activities. Speech also enjoys the distinction of being a human's highest capacity mode of communication, which makes it a particularly efficient alternative. It is unaffected by weightlessness, darkness, high levels of acceleration, and mechanical constraints, which renders it especially attractive for aerospace applications.

Due to its form, speech input contains some inherent limitations; for example, it is user dependent and sensitive to levels of background noise. Over the last few decades however, significant advances have been made in voice recognition technology, resulting in sophisticated systems that circumvent or even eliminate some of these problems. Voice recognition capabilities have progressed from single word recognition, user dependent systems, to extremely complex models that are user

independent,accept continuous speech, and use elements of artificial intelligence theory in decoding the speech signal.

In view of the above factors, it was decided to employ a voice recognition system as an aid to pilot communication in a flight simulator. The model to be used is a Verbex Series 4000 Voice Recognizer (VVR). It is speaker dependent and so it requires the user to train it with a specific vocabulary. The speaker's voice patterns are stored on a solid state cartridge and are subsequently used as templates for recognition. Two additional features of the VVR are its ability to recognize continuous speech (strings of words as opposed to single commands) and its use of structured grammars. Tests were conducted to examine the VVR's viability for use in the cockpit, to determine its reliability, and to investigate the differences in performance for different grammar structures. The results of these tests and a brief outline on the use of the VVR are presented in the following report.

2. METHOD OF USE

The procedure for using the Verbex Voice Recognizer can be outlined in a few basic steps.

1) <u>Define Grammar and Translation Table</u>

A grammar is a means of specifying the vocabulary to be used, and the format in which it is to be accepted by the VVR. A translation table defines the output messages sent to the computer when the spoken commands are recognized. Both the grammar and the translation table may be defined on the host computer using a text editor.

2) <u>Transfer to Cartridges</u>

a) Creation of the 'Master' Cartridge

Software supplied with the VVR is used to transfer the grammar and translation table to a solid state cartridge, thus creating a Master. Only a single Master need be created for each intended task.

b) Creation of the 'User' Cartridge

The VVR's training facilities enable the user to store the spoken equivalent of the vocabulary on a cartridge, thus producing a User. Any number of User cartridges may be created from a single Master. The VVR also allows retraining, so that several training sessions may be carried out to improve the voice templates.

(For more detailed information on use of the VVR see [3] and [4])

3. EQUIPMENT

The VVR is supplied as a single basic unit and requires certain additional peripherals for successful use. The extra equipment needed consists of :

1) A Host Computer

The VVR interfaces to this as an RS232-ASCII terminal, and it requires no CPU, memory, etc. . The computer used in this particular test was an IBM PC/XT.

2) An External Terminal

This acts as a means of communication with the user; for example, by prompting for speech input during training and recognition sessions. The model employed for the test was an ADM-31 terminal.

3) Solid State Cartridges

These provide storage for the voice patterns and vocabulary structures. They are usually supplied with the VVR, or may be purchased from the manufacturer.

4) Voice Planner Software

Also required is the software used for transferring the grammars to cartridge, backup of voice patterns, checking of cartridge contents, etc. This too is supplied by the manufacturer.

4. RELIABILITY TEST

This test measured the error rate of the VVR by monitoring the number of misrecognitions and rejections obtained during one test session. As low a rate as possible is desirable, and a persistent error reading of over 5% should give cause for concern. The VVR also permits the user to record several training sessions thus producing updated more accurate templates of the voice patterns. One would expect the reliability of the VVR to increase with the degree of training it has undergone. Several training passes were therefore carried out, and the trend in the error rate was monitored.

The training session used was the one supplied by the VVR, whereby when set to training mode it prompts the user to repeat specific phrases. These phrases simulate continuous speech by placing each word in the context of several other vocabulary words. The testing was carried out by setting the VVR to its recognition mode and having the subject read twice through a list of fifty phrases. Incorrect recognitions and rejections were noted manually.

The vocabulary used was a modified form of the one used in an earlier voice recognition test [1]. It provides a realistic approximation to one that could be used with the VVR in its capacity as an input to a flight simulator.
See Table 1 and Figure 1 for results of the test, and Appendix 1 for the grammar structure used.

Discussion of Results

As can be seen from Table 1, for most subjects the error rate showed a downward trend or remained constant as additional training was performed. In four out of five cases it had dropped below 5% after only two passes. The somewhat unexpected results of Subject 5 can be attributed to a greater degree of background noise than was present in the other cases.

The error values for some subjects remained more or less the same over several training passes, and this may be due to optimum recognition for that grammar being achieved after only one training session.

It was also noted that some subjects had problems with particular words, repeatedly obtaining rejections or misrecognitions from the VVR. This situation may be improved by using a feature of the VVR that allows retraining of single words. Once the troublesome areas have been isolated, they may be given a greater degree of training than the rest of the vocabulary.

An overall view of the reliability results indicates that the VVR is a viable alternative as a data entry unit, provided that a sufficient amount of training has been carried out.

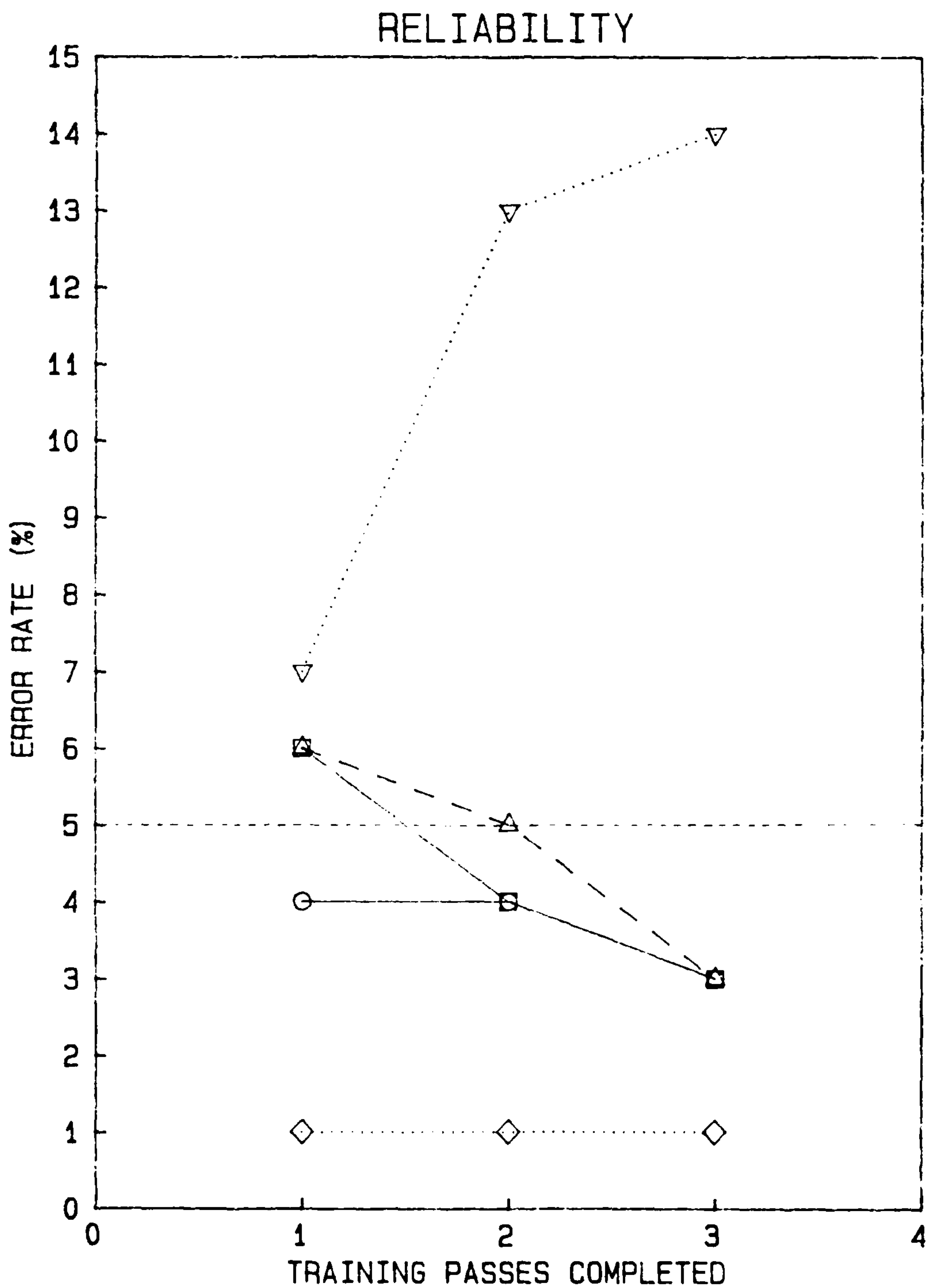


Fig. 1 - Variation in the reliability of the VVR with degree of training undergone

TABLE 1

RELIABILITY TEST RESULTS

SUBJECT	TRAINING PASSES	ERROR RATES Misrecognitions (%)	Rejections (%)	Overall Errc (%)
1	1	4	0	4
	2	3	1	4
	3	3	0	3
2	1	1	0	1
	2	1	0	1
	3	1	0	1
3	1	5	1	6
	2	2	2	4
	3	2	1	3
4	1	3	3	6
	2	3	2	5
	3	0	3	3
5	1	7	0	7
	2	13	0	13
	3	14	0	14

5. TEST ON STRUCTURED AND UNSTRUCTURED GRAMMARS

This test was run to explore the grammar definition capabilities of the VVR. A grammar may be defined with a rigid format (Appendix 1 : Structured grammar), so that the VVR accepts only specific sequences of a set of words. Grammars may also be unstructured (Appendix 2 : Unstructured grammar), when all combinations of the set of words are recognized. The former would seem to be the more reliable, for it is only allowed to accept certain combinations of words and is therefore less error prone. The latter, by being open to all combinations of words is more likely to misrecognize some of them. The unstructured grammar is, however, the more flexible of the two, for it allows a greater variety of commands to be given (due to the increased number of combinations it permits). It is also easier to use, as it does not require the speaker to remember exact sequences of words.

A test was therefore performed to monitor the relative reliabilities of the two grammars, and to investigate whether additional training could reduce the error rate of the unstructured version to acceptable levels. A single subject was used to test both the grammars under conditions similar to those of the previous reliability test. The same test list of fifty phrases was used, and retraining passes were performed until the reliabilities of both grammars reached comparable and acceptable levels. See Table 2 and Figure 2 for results and Appendices 1 and 2 for the grammars.

Discussion of Results

Initially, the error rates yielded by the unstructured grammar are seen to be significantly larger than those from the structured one. They do, however, display a downward trend as additional training is performed, yielding a value of well under 5% after seven passes.

It was also observed whilst conducting the test, that the response times obtained for the unstructured grammar were noticeably longer than those for the structured version (up to twice as long). This characteristic is to be expected, as a result of the increased number of vocabulary combinations presented to the VVR by the unstructured grammar.

It was concluded that if the magnitudes of the response times are acceptable, and if the user is prepared to perform the required number of training passes so as to significantly lower the error rate, then the unstructured grammar is a viable alternative to the structured one. Thus the flexibility of the former is gained, whilst the reliability of the latter is retained.

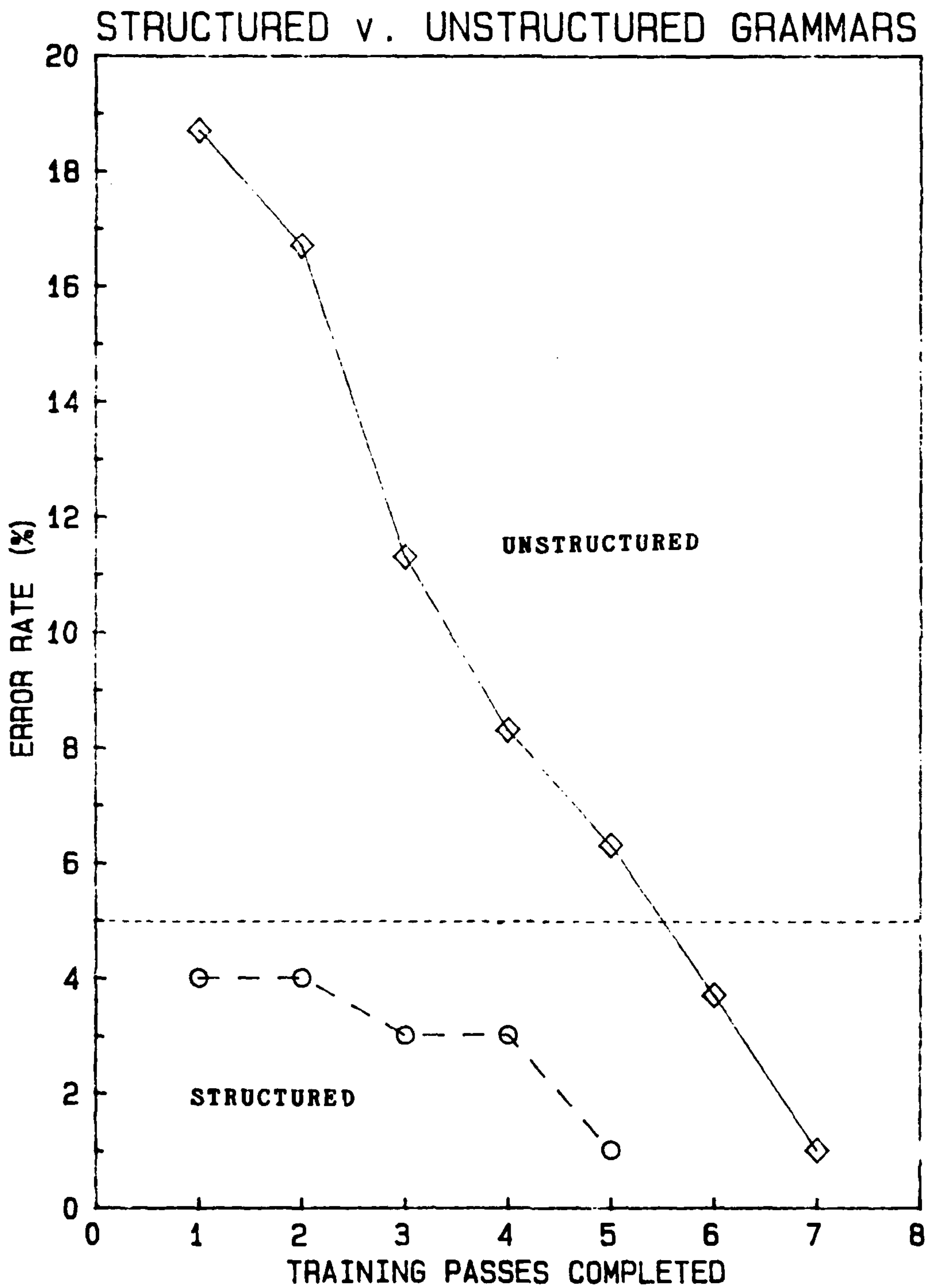


Fig. 2 - Reliability performance of the VVR for different grammar structures

TABLE 2

TEST OF STRUCTURED v. UNSTRUCTURED GRAMMARS

Results from Structured Grammar

TRAINING PASSES	ERROR RATES Misrecognitions (%)	Rejections (%)	Overall Error (%)
1	4	0	4
2	3	1	4
3	3	0	3
4	3	0	3
5	1	0	1

Results from Unstructured Grammar

TRAINING PASSES	ERROR RATES (%) RUN 1		RUN 2		RUN 3		AVERAGE		OVERALL ERROR (%)
	MR	REJ	MR	REJ	MR	REJ	MR	REJ	
1	17	2	13	7	13	4	14.3	4.3	18.7
2	18	0	16	1	13	2	15.7	1.0	16.7
3	10	3	10	1	9	1	9.7	1.7	11.3
4	6	2	9	2	5	1	6.7	1.7	8.3
5	6	0	6	1	5	1	5.7	0.7	6.3
6	2	2	2	2	3	0	2.3	1.3	3.7
7	1	0	2	0	0	0	1.0	0.0	1.0

MR = Misrecognitions REJ = Rejections

SUMMARY

The results of the reliability tests as well as the overall performance of the VVR indicate that it is suitable for such typical tasks as input to a flight simulator. Its response, however, varies with speakers, and some may be required to train it to a greater degree than others. The speed of the VVR's response changes according to the complexity of the grammar structure being tested. The impact of this characteristic should be taken into consideration if the VVR is to be used in a real time simulation.It was also noted that a structured grammar needed less training than an unstructured one to achieve the same reliability. The latter, however, may be preferred for its greater flexibility, and with sufficient training the error rates can be reduced to acceptable levels.

APPENDIX 1

Structured Grammar

This is an example of the format that a vocabulary takes when the grammar is structured.

```
over
wake-up
go-sleep
number          .DIGIT @1,5
turn            .OPTION
heading         .DIRN
execute         command        .DIGIT
change          .FACTOR        .DIGIT @1,4
altitude        .MOTION        .DIGIT @1,3        .VAL
.OUT
:erase
:abort

.DIGIT =        zero
                one
                two
                three
                four
                five
                six
                seven
                eight
                nine
                point

.OPTION =       right
                left
                on
                off

.DIRN =         northward
                southward
                eastward
                westward

.FACTOR =       frequency
                airspeed

.MOTION =       climbto
                descendto
                hold

.VAL =          thousand
                hundred

.OUT =          exit
                quit
```

APPENDIX 2

Unstructured Grammar

The grammar listed below will perform the same task as the one listed previously. This, however, is an example of a flexible or unstructured format.

```
.WORDS @1,6
.LIST @1,6
:erase
:abort

.WORDS =        execute
                command
                number
                change
                altitude
                zero
                one
                two
                three
                four
                five
                six
                seven
                eight
                nine
                point
                frequency
                airspeed
                climbto
                descendto
                hold
                thousand
                hundred

.LIST =         over
                wake-up
                go-sleep
                turn
                heading
                on
                off
                left
                right
                northward
                southward
                eastward
                westward
                exit
                quit
```

REFERENCES

[1] Huang, Chien Y., and Stengel, Robert F. "Flight and Laboratory Experiments with a Voice Recognition System", Report MAE-1754, Department of Mechanical and Aerospace Engineering, Princeton University, 1983.

[2] Wayne, Lea ed., "Trends in Speech Recognition", Prentice-Hall Inc., Englewood Cliffs, 1980.

[3] "Verbex Series 4000 Voice Planner Software User's Guide", Verbex, Exxon Corporation, Bedford, MA 01730, January, 1985.

[4] "Verbex Series 4000 Voice Recognizer User's Guide", Verbex, Exxon Corporation, Bedford, MA 01730, January, 1985.

END
DTIC
8-86